UKMED training pathway analysis

Daniel Smith

Kirsty White

Introduction

1 One of the key rationales for bringing UKMED data together is that understanding individuals' performance at different points during their study and medical career is helpful to understand the factors that make doctors more or less likely to progress through the training pathways.

2 In Phase 1 we explored whether we could develop a model that would identify the 'value-add' between entry to medical school and graduation but concluded that only limited capability existed within the current database and that a medical licensing assessment may offer greater potential to understand the learning trajectories of cohorts of students within and across medical schools.

3 Having discussed next steps with Advisory Board members, we have identified that there would be value in producing standard datasets that would enable descriptive analysis of students and doctors' movement through training pathways.

4 Two types of extract and accompanying reports have been identified:

- **a** Medical school entry profiles describing the demographics of medical school cohorts.
- **b** A workforce planning extract for those with responsibility for monitoring or planning education and training of doctors.
- **c** A summary of both extracts is described in Table 1 below and a list of fields included in **Appendix A.**

Medical school entry profiles

5 This dataset would contain demographic information from HESA on entrants to medical school including information linked on postcode such as the young participation classification (POLAR3) and UKCAT bursary information to assist track widening participation initiatives.

6 This would be used by the Medical Schools Council Selection Alliance as part of their selection monitoring work stream and by the GMC for its work on differential attainment and to enhance information provided through its *State of Medical Education and Practice* publications.

Workforce planning extract

7 This extract would contain wide-ranging data: entry profiles (from HESA and UKCAT Bursary), GMC survey census data describing doctors' specialties and training grades, Annual Review outcomes identifying doctors who do not progress and new data on where doctors work upon completion of training.

8 NHS Education for Scotland (NES) is a special NHS Board with national responsibility for education, training and workforce development for those who work in and with NHS Scotland.[1] Since 1 April 2015 Health Education England has been a Non-Departmental Public Body (NDPB) under the provisions of the Care Act 2014. Under Section 97 of this act HEE has responsibility for workforce planning[2]. Initial conversations with NES suggest that their primary interest is in understanding cross-nation flows by comparing students' country of domicile on entry to medical school to the country they go on to trainee in and finally work in.

9 The Department of Health (DH)'s role is to set overall policy and strategy direction for the health and social care system, developing evidence based policies in partnership with arm's length bodies. The Workforce Directorate in DH aims to ensure that we have the right number and mix of staff, in the right place at the right time to deliver patient care. Within that the Workforce capacity and Analytics team currently accesses rich data on an individual anonymised basis for the HCHS (Hospital and Community Health Services) workforce.

10 The workforce capacity and analytics team propose to use UKMED data to better understanding the career paths of doctors and the impact of policy and strategy, both backwards looking monitoring the impact of policy changes and forwards looking for policy development. The current model for accessing UKMED data (research applications) cannot be used by the team because they could not always commit to publishing outputs due to the confidential nature of some policy development and on occasion the need to work to very short term deadlines.

The legal framework

11 In addition to the GMC's statutory duties to set and secure standards for medical education and training, paragraph 9A(1)(b) of Part II of Schedule 1 of the 1983 Medical Act gives the GMC's a statutory duty to co-operate, in so far as is appropriate and reasonably practicable, with public bodies or other persons concerned with:

a "(i) the employment (whether or not under a contract of service) of provisionally or fully registered medical practitioners,"

b "(ii) the education or training of medical practitioners or other health care professionals"

The users of the population extract would meet this definition of public bodies we are obliged to co-operate with and the extract would be designed to allow analysis for the purposes of workforce planning.

Data that will not be included in the standard extracts

12 The following types of data that will not be included in the standard extracts:

12.1 Measures of attainment on selection tests used by medical schools (UKCAT, GAMSAT, BMAT),

12.2 Measures of attainment at school (A-levels),

12.3 Medical royal college membership exam results,

12.4 National Training Survey responses

12.5 Fitness to Practise data.

Table 1 Proposed annual extracts

Extract name	Tableau report	Description	Data sources Appendix A for field lists	Users
Medical school entry profiles	HESA Entry profiles	Extract provides demographic information on entrants to medical school including information linked on postcode such as the young participation classification (POLAR3).	HESA and UKCAT Bursary field	Data monitoring work stream for the MSC selection alliance board includes MSC data analyst and nominated medical school users. GMC analysts (beyond UKMED analysts) will have access to the de-identified dataset in the database.
Workforce Planning	To be developed. This will report on employment outcomes, for instance geographical location, area of work and type (e.g. locum) for those who have left training as defined by CCT or ARCP outcome 4.	The extract will be provided in two tables due to the differences in refresh frequency. It will be possible to link between the tables using STUDY_ID. Annual extract: VW_UKMED_HESAPROGRESSION VW_UKMED_PERSON NTS_TRAINEE This gives Demographic data from HESA on entry to medical school, including the out code[3] part of student's postcode on application to medical school. Details of where the	VW_UKMED_PERSON HESA and UKCAT Bursary NTS_TRAINEE GMC NTS census Data collected from LETBs and Deaneries ARCP_OUTCOMES ARCP outcomes collected from LETBs and Deaneries. PRACTICEHISTORY contains GMC data collected for revalidation purposes originally provided from the following payroll systems: by the four Departments of Health ESR – Electronic Staffing Records PCIS – Primary Care Information System SWISS – Scottish Workforce Information Standard System ISD Scotland's GP Contractor Database will be included	Workforce Capacity and Analytics Quarry House Department of Health Dr Louise Plewes Scottish Government - Health Workforce & Strategic Change Directorate Dr Emma Watson Planning (Medical) – Directorate of Strategy & Planning Health Education England - John Stock NHS Education for Scotland - Analysis, Intelligence and Modelling Steering Group members[6] including Dr Colin Tilley, Programme Director, Workforce Analysis Intelligence Dr Stewart Irvine, Director of Medicine

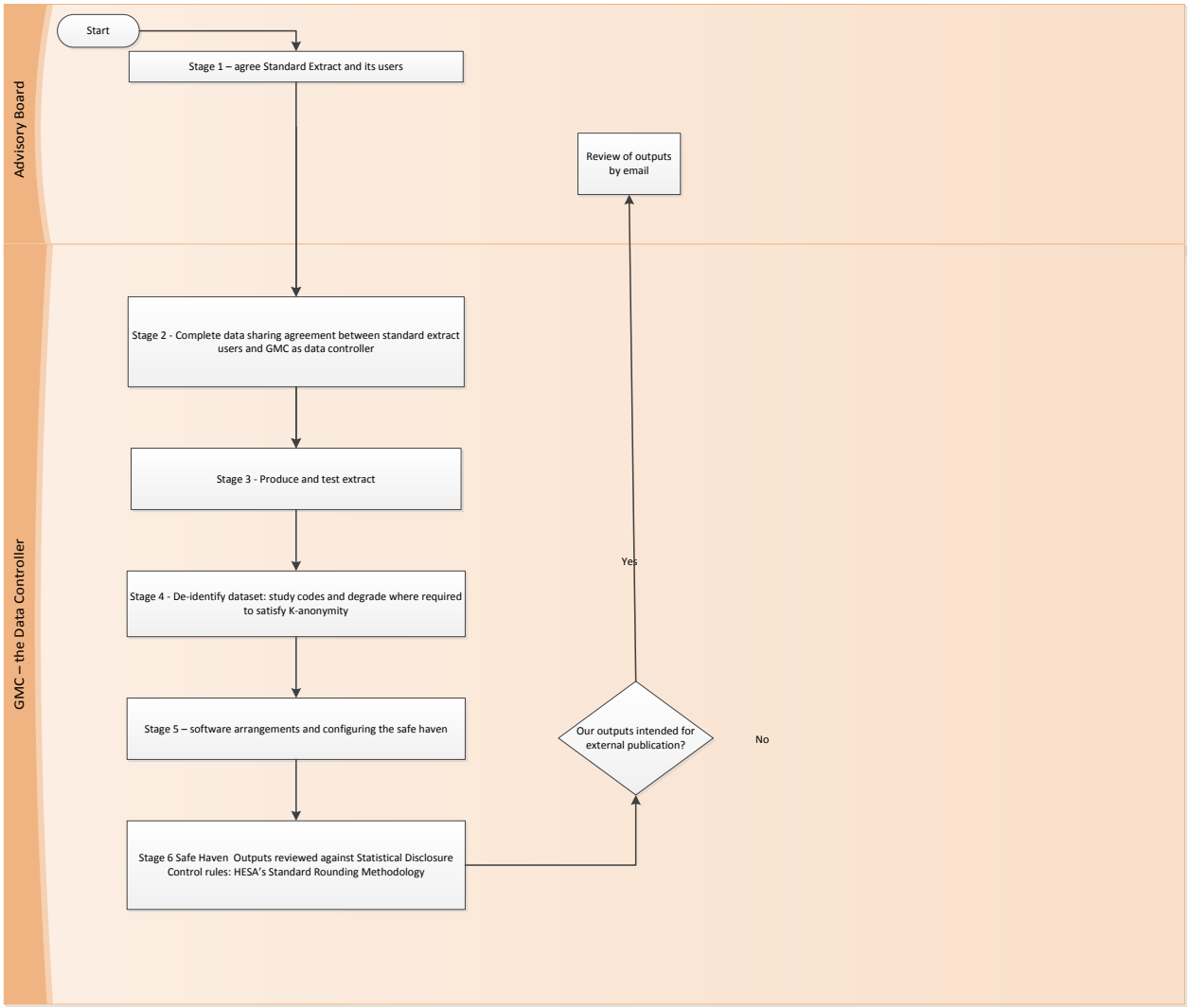
Extract name	Tableau report	Description	Data sources Appendix A for field lists	Users
		trainee was during their postgraduate training including the following: specialty, level and deanery as captured annually on the NTS Census.[4] Quarterly updates: VW_UKMED_PRACTISE_HISTORY Details of where the doctor has worked and the job role from PRACTICEHISTORY and ORGANISATION which contains data from NHS payroll systems. VW_UKMED_SPECIALTYREG Details of specialty and GP registration entries. Contains one row per instance of practice history.	Northern Ireland Business Services Organisation – this has not been refreshed recently. A process to refresh is under discussion. ORGANISATION contains the details of each ORGANISATION in PRACTICEHISTORY CR_ONS_POSTCODE contains data from National Statistics Postcode Lookup (NSPL)[5] used to assign Government Office Region and CCGs using the organisation's postcode. The view contains one row per employment assignment. VW_UKMED_SPECIALTYREG Contains data from SPECIALITIES - entries to the specialist register. PERSON which contains information on GP register entries View contains one row per specialist register entry	GMC analysts in the GMC Intelligence Unit (beyond UKMED analysts) will have access to the de-identified dataset in the database. Medical Schools Council analyst

Proposed Governance and access arrangements

13 These extracts will not be used for testing specific hypotheses and some uses of the population extract such as some governmental workforce analysis will not be published. The research process is therefore not appropriate for these extracts.

14 However, to ensure the Advisory Board has sight of all data being shared through UKMED and is able to see public uses of these data we are proposing arrangements very similar to those currently used for research extracts[7]. The same contractual constraints and requirement to use the safe haven would apply. These are summarised in figure 1 and below.

Figure 1 – Access arrangements



Stage 1 – Agree Standard Extract and its users

15 The UKMED Advisory board will have sight of the fields and definition to be included in each extract. The specification of each extract will be reviewed annually. The extracts as currently defined are in Appendix A.

Stage 2 - Complete data sharing agreement between standard extract users and GMC as data controller

16 Once the specification is finalised the GMC as Data Controller will issue a Data Sharing Agreement. This will contractually restrict the extract user's use of the data to the agreed purposes. It is important to note that the data cannot be used to support measures or decisions with respect to particular individuals, and cannot be processed in such a way that substantial damage or substantial distress is, or might be, caused to any data subject*.

Stage 3 - Produce and test extract

17 The GMC will produce the extract to the agreed specification, ensuring that the methodology of production is documented.

18 Quarterly update files for VW_UKMED_PRACTISE and VW_UKMED_SPECIALTYREG will contain the entire dataset not just the delta.

Stage 4 - De-identify dataset: study codes and degrade where required to satisfy K-anonymity

19 When providing row by row data, we will pseudonymise individual doctors. Each GMC Reference number contained within the dataset will be replaced by a unique study code. If the dataset contains multiple records with the same GMC number, these records will have the same unique study code. The unique study code will consist of a concatenation of the project code assigned on approval and a consecutive number. The GMC will hold a table that maps GMC numbers to study codes (STUDY_ID) to allow re-identification in the event of the data being queried. Study codes will only be used for one annual cycle of extracts, the same study codes will be used for tables that are refreshed quarterly during the year. This table will only be accessible to analysts working on the UKMED project. The same STUDY_ID will be used for the year and will be constant across the quarterly updates. The IDs will change when a new annual extract is issued. Old extracts will be archived and will not be available to safe haven users.

* See section 33 of the Data Protection Act (1998) here:
http://www.legislation.gov.uk/ukpga/1998/29/section/33

20 The GMC will ensure that individuals cannot be identified using a combination of demographic variables, specialty registration or employment details using data minimisation technique by applying the concept of K-anonymity. This is satisfied if K > 1 for each combination of quasi-identifiers – gender, age, medical school and so forth[*]. To achieve this it may be the case that some values will be recoded into broader categorisations. We will minimise any reduction in utility by recoding the variables least relevant to the main purpose of the report. If other techniques are used these will be outlined.

21 Data minimisation will have to consider the risks of re-identification that arise from including data in the extracts that are also publicly available, in particular the data on the List of Medical Practitioners and data on employment location. [†]

22 The GMC will maintain an archive of the extracts issued. To avoid additional complexity in satisfying K-anonymity, archived files will not be available in the safe haven. The archive is only maintained for any queries regarding outputs.

Stage 5 – software arrangements and configuring the safe haven

23 Extract users will be completing their analysis in the University of Dundee's Health Informatics Centre (HIC) Safe Haven[‡]. Users will complete a HIC/GMC Data User Agreement, which the GMC will countersign.

24 Users will need to complete a short course on Data Protection before accessing the Safe Haven and provide evidence of completion to HIC. The course "Research Data and Confidentiality" can be found at: http://byglearning.co.uk/mrcrsc-lms/course/category.php?id=1.

25 Users will be remotely logging onto a secure server located within HIC to access data and perform analysis, without being able to copy or remove the data from the secure central server.

26 The remote-access Safe Haven utilises a VMware secure environment. In this model data are no longer released externally to researchers for analysis on their own computers but placed on a server at HIC by the GMC, within a secure IT environment, where the researcher is given secure remote access to analyse it.

[*] See L. Sweeney. Achieving k-anonymity privacy protection using generalization and suppression. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10 (5), 2002; 571-588. http://dataprivacylab.org/dataprivacy/projects/kanonymity/kanonymity2.html

[†] http://www.gmc-uk.org/doctors/register/LRMP.asp

[‡] https://medicine.dundee.ac.uk/sites/medicine.dundee.ac.uk/files/Safe%20haven%20User%20Guide.pdf

Researchers will need to install the VMware client on their machine or access via http to use the safe haven[*].

27 The GMC supply the data to HIC and GMC will be responsible for all queries regarding the data. Users will have a named point of contact at the GMC for this purpose. The GMC will transfer files to HIC via a secure file transfer. Within 48 hours HIC will transfer these files to the safe haven environment (except during the 2 week Christmas/New Year period when there will be no Safe Haven support available).

28 Previously written customised code/syntax, libraries of reference data and so forth can be imported once approved by the GMC.

29 HIC are responsible for managing access to the safe haven and working with the users to ensure the required software is available. The GMC are responsible for answering any queries on the data supplied.

30 All software within the Safe Haven is licenced for academic research only, unless connected to an academic institution, it is likely that there will be an additional cost to population extract users.

31 For software that is not included as standard and where HIC Safe Haven can support it, extract users must buy the necessary licence along with the software media (to allow installation) (see table 3 for exceptions for SAS, SPSS and STATA) and pay HIC a £250 installation fee per install.

Stage 6 Safe Haven Outputs reviewed against Statistical Disclosure Control rules: HESA's Standard Rounding Methodology

32 When the user has completed their analysis, outputs intended for the public domain, for example a table of results, will be reviewed by the GMC using the following statistical disclosure controls[†]:

a 0, 1, 2 are rounded to 0

b All other numbers are rounded to the nearest multiple of 5

c Percentages based on fewer than 22.5 individuals are suppressed

d Averages based on 7 or fewer individuals are suppressed

[*] https://medicine.dundee.ac.uk/sites/medicine.dundee.ac.uk/files/Safe%20haven%20User%20Guide.pdf

[†] https://www.hesa.ac.uk/content/view/146

- The above requirements relate to headcounts, Full-Person Equivalent (FPE) and Full-Time Equivalent (FTE) data Financial data is not rounded.

33 Data output requests are processed once per day, between the hours of 9:30 and 11:30 on work-days (except during the 2 week Christmas/New Year period when there will be no Safe Haven support available). All requests made in the previous 24hrs will be processed during this period and shared with the GMC. GMC will review the files in line with statistical disclosure controls and if approved, share the output analysis files with researchers via GMC Connect within 2 working days. Researchers are strongly encouraged to leave sufficient time in their plans for their output to be reviewed before being passed to them.

34 Output intended for external consumption (for example published on an organisation's website) must be reviewed by email prior to publication. Review will be undertaken by email by persons nominated by the Advisory Board with a four week turnaround time. All external outputs must contain a clear statement on methodology, in particular criteria for inclusion in the cohorts and the details of the derivation of any variables used. Users will be expected to share derived variables with other extract users. This part of the process will be reviewed each year to ensure it is possible to resource and proportionate.

35 External publications will need to acknowledge UKMED and HESA as the data source using the following statement:

> "This report uses data from UKMED (www.ukmed.ac.uk). UKMED uses data from the Higher Education Statistics Agency Limited Source: HESA Student Record 2002/03 to 2014/15 Copyright Higher Education Statistics Agency Limited. Neither the GMC (the data controller for UKMED) or The Higher Education Statistics Agency Limited can accept responsibility for any inferences or conclusions derived by third parties from data or other information supplied by it."

Draft implementation plan

Timing	Proposed activity
April 2017	Finalise dataset fields in consultation with users Set users up in the safe haven to verify they are comfortable working in the safe haven environment

May 2017	Develop workforce planning Tableau report Refresh HESA Entry profiles report with HESA 2015/16 data and UKCAT Bursary field. Data Sharing Agreements to be signed
June 2017	Extracts to be produced and placed in the safe haven Publish a schedule for tables that will be refreshed more frequently than annually
June annually there after	Refresh the annual tables

Appendix A – Extract specification

Note that shaded fields are for internal use only to avoid re-identification.

Table: NTS_TRAINEE

This data is collected from LETBs and deaneries to administer the National Training Survey. The collection notices are here: http://www.gmc-uk.org/education/nts_documents.asp
Please see:

http://www.gmc-uk.org/NTS_2012___briefing_note_4.pdf_48235559.pdf

http://www.gmc-uk.org/NTS_2013_Briefing_Note_2.pdf_50926757.pdf

http://www.gmc-uk.org/NTS_2014___briefing_note_2___data_collection.pdf_54449847.pdf

http://www.gmc-uk.org/NTS_2015___briefing_note_2.pdf_59227197.pdf

http://www.gmc-uk.org/NTS_2016___Briefing_Note_2_FINAL_V3.pdf_63386727.pdf

Respondents validate the information supplied by the deanery/LETB. Deaneries then confirm any changes made by the survey respondent. The data provides information about the location of the trainee on the census date including their LETB/deanery, training programme and level of training.

Description last modified 06/12/2016

Columns:

NAME	DATATYPE	COMMENTS
AGORA_ID	VARCHAR2(15)	
PERSON_UID	VARCHAR2(15)	GMC doctor identifier. Known externally as GMCRefNo. Replaced by extract-specific ID in de-identified extracts.
SURVEY_YEAR	VARCHAR2(4)	The year of the survey that this record relates to.
NTS_DATA_ID	VARCHAR2(25)	Unique Siebel-generated reference ID for the NTS data

NAME	DATATYPE	COMMENTS
		record.
DEANERY_ID	VARCHAR2(15)	Trainee's Management Deanery. FK onto Organisation table.
FOUNDATION_SCHOOL	VARCHAR2(100)	The Trainee's Foundation School. Not taken from the Siebel NTS record, but derived based upon the PROGRAMME_APPROVAL_CODE (the Foundation School is stored on the PROGRAMME record. Stamped here for ease of reporting).
PROGRAMME_APPROVAL_CODE	VARCHAR2(12)	Unique reference ID for the trainee's Programme.
PROGRAMME_SPECIALTY	VARCHAR2(100)	Curriculum reference ID for the trainee's Programme
PROGRAMME_APPROVAL_CODE_2	VARCHAR2(12)	Unique reference ID for the trainee's Programme (supplied when the trainee is in a dual training programme).
PROGRAMME_SPECIALTY_2	VARCHAR2(100)	Curriculum reference ID for the trainee's second Programme (supplied if the trainee is in a dual training programme)
TRAINING_LEVEL	VARCHAR2(50)	The trainee's grade/training level.
POST_SPECIALTY	VARCHAR2(250)	Name of the trainee's Post Specialty.
POST_START_DATE	DATE	Date trainee started their post.
POST_END_DATE	DATE	Date trainee ends their post (if relevant).
BOARD_TRUST_CODE	VARCHAR2(50)	ODS code for the trainee's Board or Trust. (FK onto ORGANISATION table)
BOARD_TRUST_NAME	VARCHAR2(250)	Name of the trainee's Board or Trust. Also held on ORGANISATION table, but included here for ease of reporting.
SITE_CODE	VARCHAR2(50)	ODS Code for the trainee's Site.
SITE_NAME	VARCHAR2(250)	Name of the trainee's site. Also held in ORGANISATIONS table, but included here for ease of reporting.
IN_TRAINING_IND	VARCHAR2(1)	Indicates if the trainee is currently in a training post.
NIT_REASON	VARCHAR2(250)	Reason trainee is not in training if IN_TRAINING_IND is false.
TRAINEE_TYPE	VARCHAR2(50)	Type of trainee (e.g. LAT, FTSTA, CL, AFT, etc.)
NTN_DRN	VARCHAR2(50)	National Training Number / Deanery Reference Number.
NTN_PROGRAMME_CODE	VARCHAR2(50)	Programme code value derived from the NTN string (this value is not recorded in Siebel. A script will be executed to write this value based on the rules utilised with the ARCP validation tool).

NAME	DATATYPE	COMMENTS
SURVEY_STATUS	VARCHAR2(50)	NTS survey status Possible values: • Not started • In progress • Excluded
EXCLUSION_REASON	VARCHAR2(150)	Describes why a trainee was excluded from the survey.
SURVEY_COMPLETION_CODE	VARCHAR2(50)	Unique Siebel-generated code that is created once the trainee completes the survey.
GEOGRAPHIC_LETB	VARCHAR2(250)	Geographic LETB name (value automatically set based on the Board or Trust associated with the NTS data record)
GEOGRAPHIC_DEANERY	VARCHAR2(250)	Geographic Deanery name (value automatically set based on the Board or Trust associated with the NTS data record).
NTS_CURRICULA	VARCHAR2(100)	Referred to as "Programme Speciality Name", this is the Programme's related Curriculum Name from Siebel.

Table: VW_UKMED_PERSON

VW_UKMED_PERSON contains attributes that are considered to be about a PERSON, but particularly related to those for whom we hold student information from external data sources.

HESA and UKCAT are used to derive student attributes.

In both HESA and UKCAT, many rows can occur with the same (or differing) attributes; therefore logic has been applied in order to select the best value. As follows:

HESA:

The following values are taken from the **earliest** available Instance:

- HESA_UKPRN_FIRST
- HESA_UKPRN_FIRST_NAME
- HESA_UCAS_APP_ID
- HESA_UCAS_PERSON_ID
- HESA_PREV_INST
- HESA_PREV_INST_NAME
- HESA_DOMICILE_COUNTRY
- HESA_DOMICILE_REGION
- HESA_PARENTAL_POSTCODE
- HESA_TARIFF
- HESA_QUALENT

The following values are taken from the **latest** available Instance:

- HESA_UKPRN_LAST
- HESA_UKPRN_LAST_NAME
- HESA_OWNSTU
- HESA_XQOBTN01
- HESA_XQOBTN02

UKCAT

NOTE: For columns labelled UKCAT_* - where a person has done more than 1 UKCAT test registration, values are taken from the earliest available test registration.

Description last modified 06/12/2016

Columns:

NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15)	GMC doctor identifier. Known externally as GMCRefNo. Replaced by extract-specific ID in de-identified extracts.
BIRTH_MONTH	NUMBER(2,O)	Birth month extracted from Date of Birth held on the GMC register. Date of birth is too identifiable for inclusion in extracts.
BIRTH_YEAR	NUMBER(4,O)	Birth year extracted from Date of Birth held on the GMC register. Date of birth is too identifiable for inclusion in extracts.
GENDER	VARCHAR2 (120)	Gender
GENDER_INT	NUMBER(2,O)	Gender coded as integer for analysis: 1 = Female and 0 = Male.
PRIMARY_PMQ	VARCHAR2 (1000)	Primary medical qualification as per the GMC's register
PRIMARY_PMQ_YEAR	VARCHAR2 (16)	Primary PMQ year.
PRIMARY_PMQ_PLACE	VARCHAR2 (600)	Primary Medical Qualification awarding body
PRIMARY_PMQ_WORLD_REGION	VARCHAR2 (20)	For UKMED doctors the region will always be UK
PRIMARY_PMQ_WORLD_REGION_INT	NUMBER(2,O)	World region codes as an Integer
PRIMARY_PMQ_SHORTCODE	VARCHAR2 (200)	Code used by the GMC to identify medical schools and their awarding bodies, due to changes in awarding bodies, the same medical school can have more than one shortcode.
CURR_REG_ADD_POSTCODE_OUTCODE	VARCHAR2 (120)	Outcode, derived from CURR_REG_ADD_POSTCODE - Current postcode from the GMC register.

NAME	DATATYPE	COMMENTS
HESA_UKPRN_FIRST	VARCHAR2 (32)	The chronologically first UKPRN number that this student is found to have in HESA_ACAD_YEARS.
HESA_UKPRN_FIRST_NAME	VARCHAR2 (400)	The name of the university as found in HESA_UKPRN_FIRST. Included here for ease of reporting.
HESA_UKPRN_LAST	VARCHAR2 (32)	The chronologically last UKPRN number that this student is found to have in HESA_ACAD_YEARS.
HESA_UKPRN_LAST_NAME	VARCHAR2 (400)	The name of the university as found in HESA_UKPRN_LAST. Included here for ease of reporting.
MEDICAL_SCHOOL_FIRST	VARCHAR2 (400)	The first medical school derived from UKPRN using CR_UKPRN_NAME
MEDICAL_SCHOOL_LAST	VARCHAR2 (400)	The last medical school derived from UKPRN using CR_UKPRN_NAME
COURSE_TYPE	VARCHAR2 (120)	HESA CTITLE AND HESA COURSE_ID were manually mapped to course types in http://www.medschools.ac.uk/SiteCollectionDocuments/MSC-Entry-requirements-for-UK-medical-schools.pdf. Possible value include: Standard Entry Medicine Transferred from Oxbridge for year 3 onwards Graduate Entry Programme Medicine with a Gateway Year Medicine with a Preliminary Year Private EU Medicine
HESA_COMDATE	DATE	Commencement date of programme - The date of the student's initial commencement of studies
HESA_COMYEAR	NUMBER(4,O)	Year the student started the course extract from HESA_COMDATE
HESA_PREV_INST_DCODE	NUMBER(4,O)	*Internal use only.* The Department for Education (DfE) code for the students' previous institution. was obtained to allow linkage to school A-level results which are published by D code. Records from HESA contain codes in different formats as they migrate towards UKPRNs for all.
APSFTE_ALEVA	NUMBER(8,2)	Average point score per A level student (full-time equivalent). From DfE file linked on HESA_PREV_INST_DCODE Key Stage 5 Performance Tables - Published School Data for the year the student took A-levels. Only available for England
TALLPPE_ALEVA	NUMBER(8,2)	Average point score per A level entry. Key Stage 5 Performance Tables - Published School Data Key Stage 5 Performance Tables - Published School Data for the year the student took A-levels. Only available for England
HESA_DOMICILE_COUNTRY	VARCHAR2 (400)	Known in HESA as "Domicile - county/country".
HESA_DOMICILE_REGION	VARCHAR2 (200)	Domicile - The students' domicile prior to commencement of the course.
PARENTAL_POSTCODE	VARCHAR2 (32)	Postcode of the student at the time of their application

NAME	DATATYPE	COMMENTS
		to UCAS (therefore presumably "home" prior to university).
PARENTAL_POSTCODE_OUTCODE	VARCHAR2 (32)	#N/A
PARENTAL_POSTCODE_SRC	VARCHAR2 (20)	Set to either 'HESA' or 'UKCAT', depending on the data source used to set the PARENTAL_POSTCODE column.
HESA_TARIFF	NUMBER(8,2)	Tariff - Tariff points based on the qualifications on entry of the student. Tariff data was collected directly from UCAS prior to 2007/08 and was calculated by HESA for those entering from 2007/08 onwards. As such, tariff data will not be directly comparable across this time period.
HESA_QUALENT	VARCHAR2 (800)	Highest qualification on entry - The highest qualification that a student holds on entry. QUALENT3 was compulsory for entrants from 2010/11, from that point only existing students could be given a QUALENT2 code.
HESA_HIGHQUAL_OBTAIN	VARCHAR2 (1000)	Qualification obtained - Illustrates the qualification level achieved by the student.
UKCAT_NATIONAL_IDENTITY	VARCHAR2 (200)	Delimited list of participant national identities
UKCAT_NATIONAL_IDENTITY_INT	NUMBER(2,0)	National identify coded as an Integer.
UKCAT_REGIONAL_IDENTITY	VARCHAR2 (200)	Delimited list of participant British regional identities.
UKCAT_PARENT1GENDER	VARCHAR2 (200)	Participant parent 1 gender description
UKCAT_PARENT1OCCUPATION	VARCHAR2 (200)	participant parent 1 socio-economic class occupation description
UKCAT_PARENT1EMPLOYMENTSTATUS	VARCHAR2 (200)	Participant parent 1 socio-economic classification employment status description
UKCAT_PARENT1EMPLOYERSIZE	VARCHAR2 (200)	Participant parent 1 socio-economic classification size of employer description
UKCAT_PARENT1SUPERVISOR	VARCHAR2 (200)	Participant parent 1 socio-economic classification supervisory responsibility description
UKCAT_PARENT2GENDER	VARCHAR2 (200)	Participant parent 2 socio-economic classification gender description
UKCAT_PARENT2OCCUPATION	VARCHAR2 (200)	Participant parent 2 socio-economic classification occupation description
UKCAT_PARENT2EMPLOYMENTSTATUS	VARCHAR2 (200)	Participant parent 2 socio-economic classification employment status description
UKCAT_PARENT2EMPLOYERSIZE	VARCHAR2 (200)	Participant parent 2 socio-economic classification size of employer description
UKCAT_PARENT2SUPERVISOR	VARCHAR2 (200)	Participant parent 2 socio-economic classification supervisory responsibility description
UKCAT_DOMICILE	VARCHAR2 (200)	Area of residence of applicant.
UKCAT_NATIONALITY1	VARCHAR2 (200)	First nationality of applicant
UKCAT_NATIONALITY2	VARCHAR2 (200)	Second nationality of applicant.
ARCP_MILITARY_IND	VARCHAR2 (12)	Trainee flagged as a military trainee on at least one ARCP return
ARCP_MILITARY_IND_INT	NUMBER(2,0)	Trainee flagged as a military trainee on at least one ARCP return-
ARCP_ACADEMIC_IND	VARCHAR2 (12)	Trainee flagged as an academic trainee on at least one ARCP return.
ARCP_ACADEMIC_IND_INT	NUMBER(2,0)	Trainee flagged an academic trainee on at least one ARCP return coded as an integer.
ETHNICITY_L1	VARCHAR2 (200)	Ethnicity Level 1 information
ETHNICITY_L1_INT	NUMBER(2,0)	Ethnicity Level 1 information coded as an integer
ETHNICITY_L2	VARCHAR2 (600)	Ethnicity Level 2 information.

NAME	DATATYPE	COMMENTS
ETHNICITY_L2_INT	VARCHAR2 (48)	Ethnicity Level 2 information coded as an integer
ETHNICITY_SRC	VARCHAR2 (48)	Source system of the ETHNICITY_L1 and ETHNICITY_L2 data stored for this record.
BME	VARCHAR2 (48)	Higher level ethnicity coding: BME or white.
BME_INT	NUMBER(2,0)	#N/A Higher level ethnicity coded as an integer.
DISABILITY	VARCHAR2 (1000)	Person's disability. Source of Data held in DISABILITY_SRC.
DISABILITY _INT	#N/A	Person's disability coded as an integer
DISABILITY_SRC	VARCHAR2 (40)	Source of Person's disability as held in DISABILITY
HESA_DISABILITY_CODE_ FIRST	VARCHAR2 (48)	First disability code in the HESA data for the student
HESA_DISABILITY_LABEL_ FIRST	VARCHAR2 (48)	First disability label in the HESA data for the student
HESA_DISABILITY_CODE_ LAST	VARCHAR2 (48)	Last disability code in the HESA data for the student
HESA_DISABILITY_LABEL_ LAST	VARCHAR2 (48)	Last disability label in the HESA data for the student
LIMITED_ACTIVITIES	VARCHAR2 (1000)	Describes whether the doctor's day-to-day activities are limited because of a health problem or disability which has lasted, or is expected to last, 12 months.
LIMITED_ACTIVITIES_SRC	VARCHAR2 (40)	Source system for column LIMITED_ACTIVIITES
ADJUSTMENTS	VARCHAR2 (1000)	Describes whether the doctor required any adjustment(s) to be made so they could carry out their work in their post.
ADJUSTMENTS_SRC	VARCHAR2 (40)	Source system for column ADJUSTMENTS
UK_EDUCATED	VARCHAR2 (1000)	Describes whether the doctor completed secondary education AND their undergraduate medical degree in the UK.
SCHOOL_TYPE	VARCHAR2 (1000)	Describes which type of school the doctor mainly attended between the ages of 11 and 16. HESA State school marker
SCHOOL_TYPE_SRC	VARCHAR2 (40)	Source system for column SCHOOL_TYPE
INCOME_SUPPORT	VARCHAR2 (1000)	Describes whether the doctor's household received Income Support at any point during their school years.
INCOME_SUPPORT_SRC	VARCHAR2 (40)	Source system for column INCOME_SUPPORT
FREE_SCHOOL_MEALS	VARCHAR2 (1000)	Describes whether the doctor had free school meals.
FREE_SCHOOL_MEALS_SR C	VARCHAR2 (40)	Source system for FREE_SCHOOL_MEALS
PARENT_DEGREE	VARCHAR2 (1000)	Describes whether the doctor's parent(s) or guardian(s) completed a university degree course of equivalent.
PARENT_DEGREE_SRC	VARCHAR2 (40)	Source system for column PARENT_DEGREE
GROWING_UP_POSTCODE	VARCHAR2 (30)	Describes the postcode of the house the doctor grew up in. – self report from the trainee survey
GROWING_UP_POSTCODE _SRC	VARCHAR2 (40)	Source system for column GROWING_UP_POSTCODE – year of NTS
POLAR2	VARCHAR2 (24)	Low-participation neighbourhoods (POLAR2) - Classification of areas for young participation rates in higher education Participation rate in higher education associated with the student's domicile postcode.
POLAR2_SRC	VARCHAR2 (40)	Source system for column POLAR2
POLAR3	VARCHAR2 (24)	Low-participation neighbourhoods (POLAR3) - Classification of areas for young participation rates

NAME	DATATYPE	COMMENTS
		in higher education .Participation rate in higher education associated with the student's domicile postcode.
POLAR3_SRC	VARCHAR2 (40)	Source system for column POLAR3
SEC	VARCHAR2 (12)	Socio-economic classification Socio-economic classification of the student's parent (if under 21) or the student themselves when over 21. Mapping table here: https://www.hesa.ac.uk/index.php?option=com_studrec&task=show_file&mnl=13051&href=a^_^SEC.html
SEC_INT	NUMBER(2,0)	SEC coded as an Integer
SEC_SRC	VARCHAR2 (40)	Source system for column SEC
NSSEC	VARCHAR2 (400)	National Statistics socio-economic five-point scale classification 1 = managerial and professional occupations 2 = intermediate occupations 3 = small employers and own account workers 4 = lower supervisory and technical occupations 5 = semi-routine and routine occupations
NSSEC _INT	NUMBER(2,0)	NSSEC coded as an integer
NSSEC_SRC	VARCHAR2 (40)	Source of NSSEC
SOC2000	VARCHAR2 (48)	Occupation code - SOC code of student if ages 21 or over at the start of course, or the parents SOC code if under 21 Occupational code of the student's parent (where the student is under 21), OR the occupation of the student themselves when they start training at over 21.
SOC2000_SRC	VARCHAR2 (40)	Source system for column SOC2000.
PARED	VARCHAR2 (12)	Parental education records whether a student's parents had higher education qualifications. HESA provide further details of on how the data are collected here: https://www.hesa.ac.uk/component/studrec/show_file/14051/a%5E_%5EPARED.html
PARED _INT	NUMBER(2,0)	PARED coded as an Integer
PARED_SRC	VARCHAR2 (40)	Source column for column PARED
QUINTILE	VARCHAR2 (200)	Rank of IMD zone within country where 1 = MOST deprived.-Note that they are constructed differently in each country. http://www.neighbourhood.statistics.gov.uk/dissemination/Info.do?m=0&s=1460015515128&enc=1&page=analysisandguidance/analysisarticles/indices-of-deprivation.htm&nsjs=true&nsck=false&nssvg=false&nswid=1366
COUNTRY	VARCHAR2 (400)	Country where school is located
DFEID	VARCHAR2 (50)	Schools code in UKCAT data Internal use only
CATEGORYID	VARCHAR2 (50)	UCAS school category identifier 1 = UK School

NAME	DATATYPE	COMMENTS
		2 = Under Sixteen School 3 = Further Education College 4 = Southern Irish School 5 = EU School 6 = Non-EU School 7 = Overseas School `` 8 = unknown
SUBCATEGORYID	VARCHAR2 (50)	1 = agricultural and horticultural college 2 = art, design and performing arts 3 = comprehensive school 4 = further education 5 = grammar school 6 = grant maintained (special school) 7 = grant maintained secondary school (state) 8 = higher education 9 = independent school 10 = language school 11 = sixth form centre 12 = sixth form college 13 = special school 14 = technical college 15 = tertiary college 16 = other secondary school 17 = other school 18 = not applicable 19 = unknown
STATUS	VARCHAR2 (8)	Open/closed status of school
CATEGORYDESCRIPTION,	VARCHAR2 (50)	Label for CATEGORYID
SUBCATEGORYDESCRIPTION	VARCHAR2 (200)	Label for SUBCATEGORYID
QYPR	NUMBER(2,0)	Young participation quintile 1(low)-5(high) and (unclassified) The young participation classification (POLAR3) is based the participation in HE of young people who reached 18 between 2005-2009. . Taken from http://www.hefce.ac.uk/analysis/yp/POLAR/POLAR3,data/
QAHE	NUMBER(2,0)	Adult HE qualification quintile 1(low)-5(high) and (unclassified) The adult qualification level classification is based on the proportion of people aged 16-74 with HE qualifications from the 2001 Census Key Statistics Table 13 Taken from http://www.hefce.ac.uk/analysis/yp/POLAR/POLAR3,data/
IDACI_SCORE	NUMBER(4,2)	Income Deprivation Affecting Children Index (IDACI) constructed by the Social Disadvantage Research Centre at the University of Oxford as part of the Indices of Deprivation 2007. All figures can only be reproduced if the source (Department of Communities and Local Government, Indices of Deprivation 2007) is fully acknowledged. Please Note All scores are presented to two decimal places. In some cases LSOAs with apparently the same score

NAME	DATATYPE	COMMENTS
		will be given different ranks, according to their actual score.
RANK_OF_IDACI	NUMBER(4,0)	The SOA with a rank of 1 is the most deprived, and 32482 the least deprived, on this overall measure.
IDAOPI_SCORE	NUMBER(4,2)	Income Deprivation Affecting Older People Index constructed by the Social Disadvantage Research Centre at the University of Oxford as part of the Indices of Deprivation 2007. All figures can only be reproduced if the source (Department of Communities and Local Government, Indices of Deprivation 2007) is fully acknowledged. Please Note All scores are presented to two decimal places. In some cases LSOAs with apparently the same score will be given different ranks, according to their actual score.
RANK_OF_IDAOPI	NUMBER(4,0)	The SOA with a rank of 1 is the most deprived, and 32482 the least deprived, on this overall measure.
GRADUATEONENTRY	VARCHAR2 (50)	Based on HESA_QUALENT, the following values are defined as graduate entry 'First degree of UK institution', 'First degree with honours leading to Qualified Teacher Status (QTS)/registration with a General Teaching Council (GTC)', 'Higher degree of UK institution', 'Non-UK doctorate degree', 'Non-UK first degree', 'Non-UK masters degree', 'PGCE with QTS/GTC registration', 'PGCE without QTS/GTC registration', 'Postgraduate diploma or certificate, excluding PGCE', 'UK doctorate degree', 'UK first degree with honours', 'UK masters degree', 'UK ordinary (non-honours) first degree', 'Postgraduate Certificate in Education or Professional Graduate Diploma in Education', 'Graduate of other overseas institution', 'Professional Graduate Certificate in Education', 'Integrated undergraduate/postgraduate taught masters degree on the enhanced/extended pattern', 'Graduate of EU institution'
QUINTILECLEANED	NUMBER(4,0)	Value not 1through to 5 removed
QUINTILECLEANEDLABEL	#N/A	Quintile with label for clarity '1 - Least deprived' through to '5 - Most deprived'
QUINTILENONGRADUATEENTRY	NUMBER(4,0)	IMD quintile for non-graduate only as postcode on entry more likely to be the parental postcode.
SEC_COMBINED	VARCHAR2 (50)	Uses the value from HESA if present; if not present uses the value from the UKCAT registration form.
BURSARY	VARCHAR2 (50)	Flag to indicate presence in the UKCAT Bursary table

Table: VW_UKMED_HESAPROGRESSION

VW_UKMED_HESAPROGRESSION contains one row per person with the most recent RSEND value that is not 'Unknown' from HESA_ACAD_YEARS.

NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15)	GMC doctor identifier. Known externally as GMCRefNo. Replaced by extract-specific ID in de-identified extracts.
HESA_ACYEAR	VARCHAR2(7)	Academic year.
HESA_RSNEND	VARCHAR2(50)	Reason for leaving - Indicates for what reason the student left medical school See: https://www.hesa.ac.uk/collection/c16051/a/rsnend/.html/.html/
HESA_ENDDATE	DATE	The date the student left the student instance detailed in the given record. See: https://www.hesa.ac.uk/collection/c15051/a/enddate/.html/.html/
HESA_YEAR_PRG	VARCHAR2(50)	This field indicates the year number of the course that the student is currently studying. This could be different from the year of student if the student has changed course or re-taken a year. See: https://www.hesa.ac.uk/collection/c16051/a/yearprg/.html/.html/
HESA_YEAR_STUDY	VARCHAR2(50)	This field indicates the year number that the student is in since enrolling for a course leading to the student's qualification aim (whether or not the intended subject or class has changed) i.e. number of years on this student instance. This could be different from the year of course if the student has changed course or retaken a yea See: https://www.hesa.ac.uk/collection/c15051/a/yearstu/.html/.html/

Table: VW_UKMED_PRACTICEHISTORY

Contains data from

PRACTICEHISTORY

ORGANISATION

These tables contain data from

ESR – Electronic Staffing Records (refreshed at least quarterly)

PCIS – Primary Care Information System (refreshed at least quarterly)

SWISS – Scottish Workforce Information Standard System (refreshed at least quarterly)

Northern Ireland Business Services Organisation. Note that very little data are available from Northern Ireland. Data has not been refreshed recently.

There is one row per instance of practice history – i.e. employment episode

One record should be provided for every "assignment" that the doctor has with a given employer. An assignment should be uniquely defined as the combination of the following data items:

- GMC Reference Number (Doctor UID)
- Site code
- Employing (or Parent) Organisation Code
- Employment start date

NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15)	GMC doctor identifier. Known externally as GMCRefNo. Replaced by extract specific id in de-identified research extracts.
AREA_OF_WORK	VARCHAR2	Source table: PRACTICEHISTORY. This is the area, function or specialty where the work activity takes place. See National Workforce Data Set (NWD) - http://content.digital.nhs.uk/datasets/nwd
ASSIGNMENT_CATEGORY	VARCHAR2	Source table: PRACTICEHISTORY Data are only available for ESR contains information on whether the person is active in the role
DATA_SOURCE	VARCHAR2	Source table: PRACTICEHISTORY. Gives the data source of the record, one of the following: ESR –

NAME	DATATYPE	COMMENTS
		Electronic Staffing Records PCIS – Primary Care Information System SWISS – Scottish Workforce Information Standard System Northern Ireland Business Services Organisation
SPECIALITY_AREA	VARCHAR2	Source table: PRACTICEHISTORY. ESR only, gives the Doctor's specialty area of work, more detailed than area of work. Some values are CCT specialty values.
START_DATE	DATE	Source table: PRACTICEHISTORY
END_DATE	DATE	Source table: PRACTICEHISTORY
WORK_PATTERN	VARCHAR2	Source table: PRACTICEHISTORY Full-time or part-time This is not available for Northern Ireland cases
GRADE	VARCHAR2	Source table: PRACTICEHISTORY. ESR only, this contains codes for NHS payscales and will allow identification of those working at consultant level versus those working as Associate Specialists. Staff Grade or locums etc. Further details are available here - http://www.nhsemployers.org/~/media/Employers/Documents/Pay%20and%20reward/Pay%20and%20Conditions%20Circular%20MD%2012015.pdf
JOB_ROLE	VARCHAR2	Source table: PRACTICEHISTORY. Identifies GP Locum GP Registrar General Practitioner
JOB_ROLE_DTL	VARCHAR2	Source table: PRACTICEHISTORY. Include information on the types of GP from PCIS.
PRACTICE_TYPE	VARCHAR2	Source table: PRACTICEHISTORY Identifies locum and permanent contract types.
DB_IND	VARCHAR2	ORGANISATION
ODS_CODE_SITE	VARCHAR2	ORGANISATION. NHS side code
ORG_NAME_SITE	VARCHAR2	ORGANISATION
ODS_CODE_ORG	VARCHAR2	ORGANISATION. NHS organisation code
ORG_NAME_ORG	VARCHAR2	ORGANISATION
ODS_PAR_NAME_ORG	VARCHAR2	ORGANISATION
GOR_ORG_POSTCODE	VARCHAR2	Government office region from Group from National Statistics Postcode Lookup (NSPL) in CR_ONS_POSTCODE see National Statistics Postcode Lookup (February 2017) User Guide. Available from http://geoportal.statistics.gov.uk/ Office for National Statistics (Edition:February 2017) National Statistics Postcode Lookup User Guide Available from: https://www.arcgis.com/sharing/rest/content/items/8e409cfe4d0a4971986343f3919021e3/data The region code for each postcode. Pseudo codes are included for Wales, Scotland, Northern Ireland, Channel Island and Isle of Man
CCG_ORG_POSTCODE	VARCHAR2	Clinical Commissioning Group from National Statistics Postcode Lookup (NSPL) in CR_ONS_POSTCODE see

NAME	DATATYPE	COMMENTS
		National Statistics Postcode Lookup (February 2017) User Guide. Available from http://geoportal.statistics.gov.uk/ Office for National Statistics (Edition:February 2017) National Statistics Postcode Lookup User Guide Available from: https://www.arcgis.com/sharing/rest/content/items/8e409cfe4d0a4971986343f3919021e3/data

Table: VW_UKMED_SPECIALTYREG

Contains data from

SPECIALITIES - entries to the specialist. This contains data that is found on the publically available list of medical practitioner here - http://www.gmc-uk.org/doctors/register/LRMP.asp. To guard against re-identification of cases, dates are set to year only.

PERSON which contains information on GP register entries

It contains one row per specialist register entry

NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(15)	GMC doctor identifier. Known externally as GMCRefNo. Replaced by extract specific id in de-identified research extracts.
CURR_INCOME_DISCOUNT_IND	VARCHAR2	PERSON
CURR_LIC_STATUS_IND	VARCHAR2	PERSON
CURR_REG	VARCHAR2	PERSON
GP_IND	VARCHAR2	PERSON
YEARGPREGISTER	#N/A	The year the person joined the GP register.
NATIONALITY	VARCHAR2	PERSON
YEARFIRSTSPECIALTYENTRY	#N/A	The year of first specialty registration
SPECIALIST_IND	VARCHAR2	PERSON
YEARSPECIALTYENTRY	#N/A	The year of entry to the specialist register for this specialty.
ROYAL_COLLEGE	VARCHAR2	SPECIALTIES
ROYAL_COLLEGE_NAME	VARCHAR2	SPECIALTIES
SPECIALTY	VARCHAR2	SPECIALTIES
SPECIALTY_GROUP	VARCHAR2	SPECIALTIES
SUB_SPECIALTY	VARCHAR2	SPECIALTIES

Table: ARCP_OUTCOMES

Each year, Local Education Training Boards (LETBs) and deaneries provide the GMC with ARCP (Annual Review of Competence Progression) data, which describes trainees' progression (or not) through their medical training.

A trainee may have more than one ARCP record in a return file, as an outcome is awarded for each specialty undertaken by the trainee. A trainee may also have more than one ARCP event per year (e.g. some core programmes rotate every 8 months, so a trainee could potentially have 2 ARCPs in a given reporting year). Some trainees receive an outcome 5 (insufficient evidence); followed by another outcome upon presentation of the evidence in the same year.

The data collection notices that the GMC issues for this data collection are available here: http://www.gmc-uk.org/education/23877.asp

Columns:

NAME	DATATYPE	COMMENTS
PERSON_UID	VARCHAR2(12)	GMC doctor identifier. Known externally as GMCRefNo. Replaced by extract-specific ID in de-identified research extracts.
ARCP_EVENT_ID	VARCHAR2(50)	Unique ID for each record in the ARCP return file. The ID will be generated by the GMC after the files are returned to the GMC and will be used to join to records in the ARCP_REASONS table.
ARCP_SUBMITTING_DEANERY	VARCHAR2(100)	The deanery that returned the ARCP record to the GMC.
ARCP_DATA_YEAR	VARCHAR2(4)	The year the data was returned to the GMC.
ARCP_REVIEW_DATE	DATE	Date of RITA or ARCP event.
ARCP_PERIOD_START_DATE	DATE	Start of the period that assessment refers to.
ARCP_PERIOD_END_DATE	DATE	End of the period that the assessment refers to.
ARCP_LEVEL_ASSESSED	VARCHAR2(50)	Grade of trainee at the point of assessment, or at the point of the missed assessment.
ARCP_POST_TYPE	VARCHAR2(50)	Trainee's Post type (e.g. Core, Specialty, Foundation)
ARCP_MILITARY_IND	VARCHAR2(3)	Indicates if the trainee was a defence DPMD trainee (at the time of the ARCP/RITA).
ARCP_ACADEMIC_IND	VARCHAR2(3)	Indicates if the trainee was an academic trainee (at the time of the ARCP/RITA).
ARCP_REMAIN_ON_ACADEMIC_PROG	VARCHAR2(250)	List of value that describes if (and why) the trainee will remain on their academic programme.
ARCP_SPECIALTY	VARCHAR2(250)	Specialty to which the ARCP/RITA refers.
ARCP_NOT_FULL_TIME_IND	VARCHAR2(3)	Indicates if the trainee was less than full time during any part of the period covered by the ARCP.
ARCP_OUTCOME_CODE	VARCHAR2(250)	ARCP / RITA Outcome. Possible values; 1; 2; 3; 4; 5; 6; 7.1; 7.2; 7.3; 7.4; 8; 9; C; D; E; F; G
ARCP_OUTCOME_CLASSIF	VARCHAR2(50)	Please see:

NAME	DATATYPE	COMMENTS
		http://www.gmc-uk.org/education/23861.asp
ARCP_OUTCOME_DESCRIPTION	VARCHAR2(250)	Description for OUTCOME_CODE – see http://www.gmc-uk.org/education/23861.asp
ARCP_OUTCOME_APPL_TO_FOUND	VARCHAR2(50)	Please see: http://www.gmc-uk.org/education/23861.asp
ARCP_OUTCOME_TYPE	VARCHAR2(50)	Describes if this is a clinical or academic outcome.
ARCP_BENCHMARK_GROUP	VARCHAR2(50)	Derived variable for grouping specialties for reporting purposes.
ARCP_OUTCOME_ORDERED	NUMBER(8,2)	Outcomes on an ordinal scale '1'=1 '2'=3 '3'=4 '4'=4 '5'=2 '6'=1 '7'=1 '7.1'=1 '7.2'=3 '7.3'=4 '7.4'=2 '8'=100 '9'=100 'C'=1 'D'=3 'E'=4 'F'=100 'G'=1 1 'Satisfactory progression' 2 'Insufficient evidence presented' 3 'Targeted training required (but training time not extended)' 4 'Extended training time required/left programme' 100 'Out of programme' The outcomes awarded were coded into an ordinal scale following the method developed by Tiffin et al (2014). Tiffin, P. A., Illing, J., Kasim, A. S. & McLachlan, J. C. (2014). Annual Review of Competence Progression (ARCP) performance of doctors who passed Professional and Linguistic Assessments Board (PLAB) tests compared with UK medical graduates: national data linkage study. BMJ 348: g2622. Available at: http://www.bmj.com/content/348/bmj.g2622
ARCP_OUTCOME_ORDERED_DES	VARCHAR2(100)	Labels for ordered number

End notes

[1] NHS Education for Scotland (2014) *A refreshed strategic framework for 2014-19* Available at:

[2] http://www.legislation.gov.uk/ukpga/2014/23/section/97/enacted

[3] http://www.bph-postcodes.co.uk/guidetopc.cgi

[4] Example from 2016- http://www.gmc-uk.org/NTS_2016___Briefing_Note_2_FINAL_V3.pdf_63386727.pdf there is a similar collection notice for each year.

[5] Available from http://geoportal.statistics.gov.uk/
Office for National Statistics (Edition:February 2017) *National Statistics Postcode Lookup User Guide* Available from:
https://www.arcgis.com/sharing/rest/content/items/8e409cfe4d0a4971986343f3919021e3/data

[6] http://www.nes.scot.nhs.uk/education-and-training/by-theme-initiative/analysis,-intelligence-and-modelling/aim-steering-group.aspx

[7] UK Medical Education Database (UKMED). *UKMED Process for completing UKMED Research v2 December 2016* Available at: http://www.ukmed.ac.uk/documents/UKMED_research_process.pdf [Accessed 28 December 2016]